18. Web Security and Attacks (Part 2)

Blase Ur and David Cash
February 17th, 2023
CMSC 23200 / 33250

THE UNIVERSITY OF
CHICAGO

CSRF

Cross-Site Request Forgery (CSRF)

- Goal: Make a user perform some action on a website without their knowledge
  - Trick the browser into having them do this
- Main idea: Cause a user who's logged into that website to send a request that has lasting effects

Cross-Site Request Forgery (CSRF)

- Prerequisites:
  - *Victim* is logged into *important.com* in a particular browser
  - *important.com* accepts GET and/or POST requests for important actions
  - *Victim* encounters *attacker's* code in that same browser

CSRF Example

- *Victim* logs into *important.com* and they stay logged in (within some browser)
  - Likely an auth token is stored in a cookie
- *Attacker* causes *victim* to load `https://www.important.com/transfer.php?amount=100000000&recipient=blase`
  - This is a GET request. For POST requests, auto-submit a form using JavaScript
- Transfer money, cast a vote, change a password, change some setting, etc.

CSRF: How?!

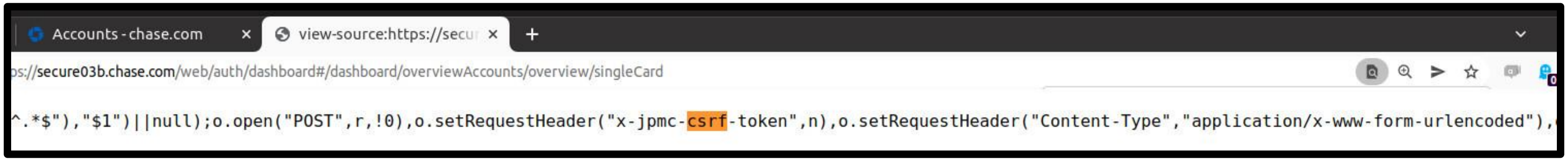
- On *blaseur.com* have <a href=“`URL`”>Cat photos</a>
- Send an HTML-formatted email with <img src=“`URL`”>
- Have a hidden form on *blaseur.com* with JavaScript that submits it when page loads
- Etc.

CSRF: Why Does This Work?

- Recall: Cookies for *important.com* are automatically sent as HTTP headers with every HTTP request to *important.com*
- *Victim* doesn’t need to visit the site explicitly, but their browser just needs to send an HTTP request
- Basically, the browser is confused
  - “Confused deputy” attack

CSRF: Key Mitigations

- Check HTTP referrer *(less good)*
  - Can sometimes be forged
- CSRF token *(standard practice)*
  - “Randomized” value known to *important.com* and inserted as a hidden field into forms
  - Key: not sent as a cookie, but sent as part of the request (HTTP header, form field, etc.)


Accounts - chase.com
view-source:https://secu
ps://secure03b.chase.com/web/auth/dashboard#/dashboard/overviewAccounts/overview/singleCard

```
^.*$"),"$1")||null);o.open("POST",r,!0),o.setRequestHeader("x-jpmc-csrf-token",n),o.setRequestHeader("Content-Type","application/x-www-form-urlencoded"),
```

XSS

Cross-Site Scripting (XSS)

- Goal: Run JavaScript on someone else’s domain to access that domain’s DOM
  - If the JavaScript is inserted into a page on *victim.com* or is an external script loaded by a page on *victim.com*, it follows *victim.com*’s same origin policy
- Main idea: Inject code through either URL parameters or user-created parts of a page

Cross-Site Scripting (XSS)

- Variants:
  - *Reflected XSS*: The JavaScript is there only temporarily (e.g., search query that shows up on the page or text that is echoed)
  - *Stored XSS:* The JavaScript stays there for all other users (e.g., comment section)
- Prerequisites:
  - HTML isn't (completely) stripped
  - *victim.com* echoes text on the page
  - *victim.com* allows comments, profiles, etc.

XSS: How?

- Type *<script>*`EVIL CODE();`*</script>* into form field that is repeated on the page
- Do the same, but as a URL parameter
- Add a comment (or profile page, etc.) that contains the malicious script
- Malicious script accesses sensitive parts of the DOM (financial info, cookies, etc.)
  - Change some values
  - Exfiltrate info (load *attacker.com/?q=SECRET*)

XSS: Why Does This Work?

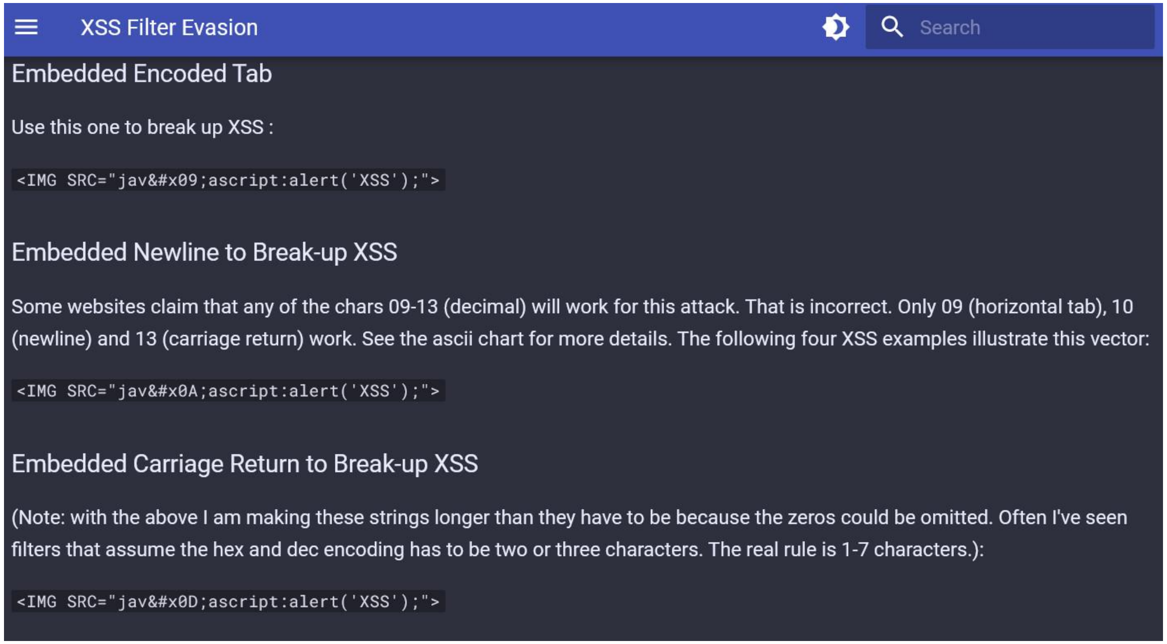
- All scripts on *victim.com* (or loaded from an external source by *victim.com*) are run with *victim.com* as the origin
  - By the Same Origin Policy, can access DOM

XSS: Key Mitigations

- Sanitize / escape user input
  - Harder than you think!
  - Different encodings
  - <img onmouseover=“`EVIL CODE();`” />
  - Use libraries to do this!
- Define Content Security Policies (CSP)
  - Specify where content (scripts, images, media files, etc.) can be loaded from
  - `Content-Security-Policy: default-src 'self' *.trusted.com`

XSS: Subtleties

- See https://cheatsheetseries.owasp.org/cheatsheets/XSS_Filter_Evasion_Cheat_Sheet.html for lots of examples of trying to evade filters



Embedded Encoded Tab

Use this one to break up XSS :

```
<IMG SRC="jav&#x09;ascript:alert('XSS');">
```

Embedded Newline to Break-up XSS

Some websites claim that any of the chars 09-13 (decimal) will work for this attack. That is incorrect. Only 09 (horizontal tab), 10 (newline) and 13 (carriage return) work. See the ascii chart for more details. The following four XSS examples illustrate this vector:

```
<IMG SRC="jav&#x0A;ascript:alert('XSS');">
```

Embedded Carriage Return to Break-up XSS

(Note: with the above I am making these strings longer than they have to be because the zeros could be omitted. Often I've seen filters that assume the hex and dec encoding has to be two or three characters. The real rule is 1-7 characters.):

```
<IMG SRC="jav&#x0D;ascript:alert('XSS');">
```

SQL Injection

Very Basic MySQL

- Goal: Manage a database on the server
- Create a database:
  - `CREATE DATABASE cs232;`
- Delete a database:
  - `DROP DATABASE cs232;`
- Use a database (subsequent commands apply to this database):
  - `USE cs232;`

Very Basic MySQL

- Create a table:
  - `CREATE TABLE potluck (id INT NOT NULL PRIMARY KEY AUTO_INCREMENT, name VARCHAR(20), food VARCHAR(30), confirmed CHAR(1), signup_date DATE);`
- See your tables:
  - `SHOW TABLES;`
- See detail about your table:
  - `DESCRIBE cs232;`

Very Basic MySQL

- Insert data into a table:

  -
    ```
    INSERT INTO potluck (id, name,
    food, confirmed, signup_date)
    VALUES (NULL, 'David Cash', 'Vegan
    Pizza', 'Y', '2022-02-18’);
    ```
- Edit rows of your table:

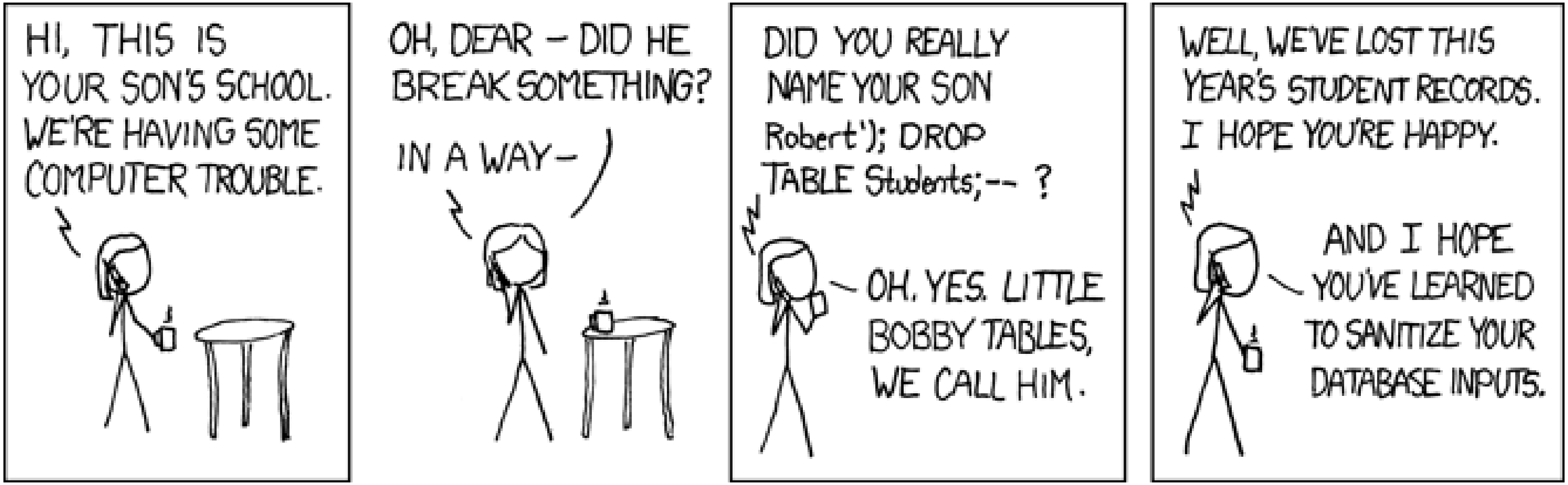
  -
    ```
    UPDATE potluck SET food = 'None'
    WHERE name = 'David Cash';
    ```
- Get your data:

  -
    ```
    SELECT * FROM potluck;
    ```

SQL Injection

- Goal: Change or exfiltrate info from *victim.com*’s database
- Main idea: Inject code through the parts of a query that you define

SQL Injection


HI, THIS IS YOUR SON'S SCHOOL. WE'RE HAVING SOME COMPUTER TROUBLE.
OH, DEAR – DID HE BREAK SOMETHING?
IN A WAY–
DID YOU REALLY NAME YOUR SON Robert'); DROP TABLE Students;-- ?
OH. YES. LITTLE BOBBY TABLES, WE CALL HIM.
WELL, WE'VE LOST THIS YEAR'S STUDENT RECORDS. I HOPE YOU'RE HAPPY.
AND I HOPE YOU'VE LEARNED TO SANITIZE YOUR DATABASE INPUTS.

SQL Injection

- Prerequisites:
  - Victim site uses a database
  - Some user-provided input is used as part of a database query
  - DB-specific characters aren't (completely) stripped

SQL Injection: How?

- Enter DB logic as part of query you impact
- Back-end query
  - `SELECT * FROM USERS WHERE USER='' AND PASS='';`
- For password of user blase , attacker gives:
  - `' OR '1'='1`
- Straightforward insertion:
  - `SELECT * FROM USERS WHERE USER='blase' AND PASS='' OR '1'='1';`

SQL Injection: Why Does This Work?

- Database does what you ask in queries!

SQL Injection: Key Mitigations

- Sanitize / escape user input
  - Harder than you think!
  - Different encodings
  - Use libraries to do this!
- **Prepared statements** from libraries handle escaping for you!
- Use PHP's mysqli (in place of mysql) with prepared statements
  - https://www.w3schools.com/php/php_mysql_prepared_statements.asp

Additional Web Topics

Processing Data on the Server

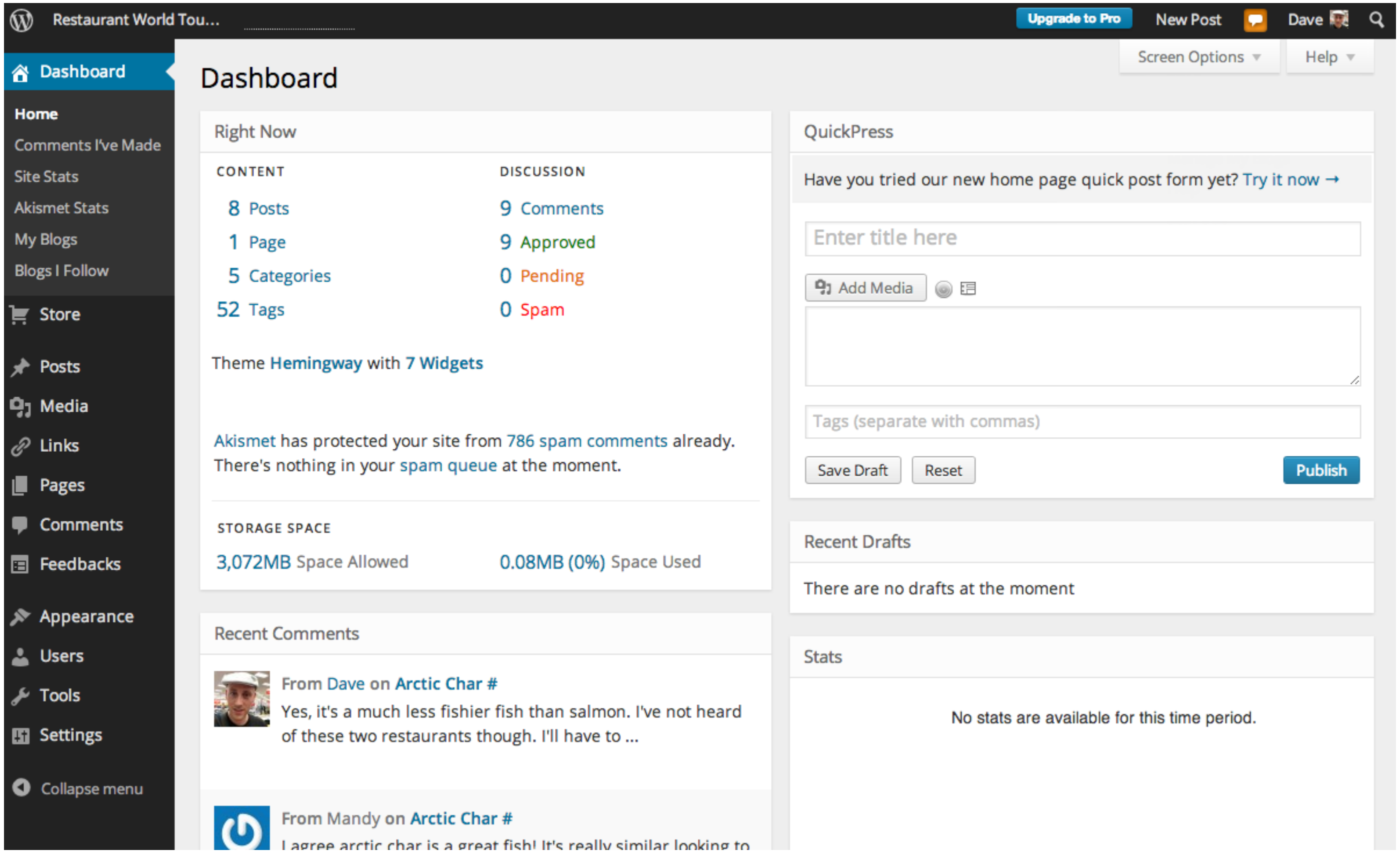
- JavaScript is client-side
- Server-side you find Perl (CGI), PHP, Python (Django)
- Process data on the server
- What happens if this code crashes?

Storing Data on the Server

- Run a database on the server
- MySQL, SQLite, MongoDB, Redis, etc.
- You probably don’t want to allow access from anything other than *localhost*
- You definitely don’t want human-memorable passwords for these

CMS (Content Management System)

- WordPress (PHP + MySQL), Drupal



Dashboard

Right Now

CONTENT

8 Posts
1 Page
5 Categories
52 Tags

DISCUSSION

9 Comments
9 Approved
0 Pending
0 Spam

Theme Hemingway with 7 Widgets

Akismet has protected your site from 786 spam comments already. There's nothing in your spam queue at the moment.

STORAGE SPACE

3,072MB Space Allowed
0.08MB (0%) Space Used

Recent Comments

From Dave on Arctic Char #
Yes, it's a much less fishier fish than salmon. I've not heard of these two restaurants though. I'll have to ...

From Mandy on Arctic Char #
I agree arctic char is a great fish! It's really similar looking to

QuickPress

Have you tried our new home page quick post form yet? Try it now →

Enter title here

Tags (separate with commas)

Recent Drafts

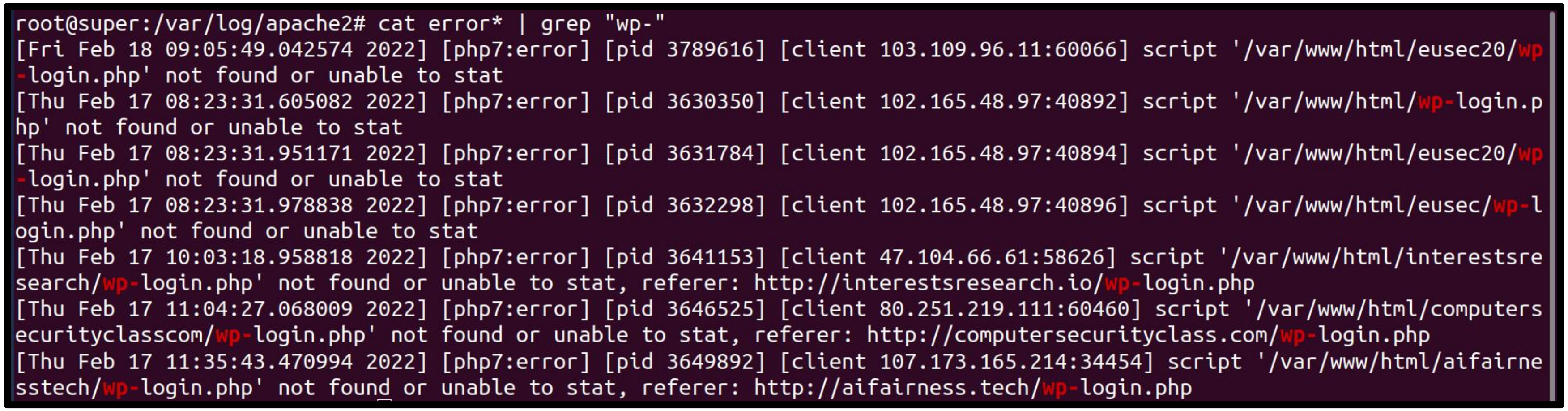
There are no drafts at the moment

Stats

No stats are available for this time period.

CMS Defaults / Vulnerabilities

- WordPress attempted logins:

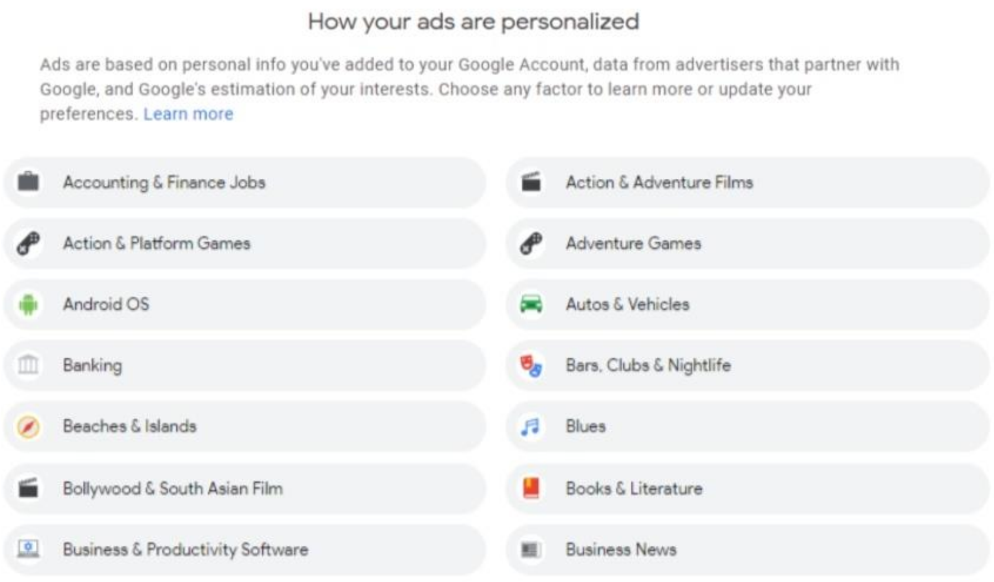


```
root@super:/var/log/apache2# cat error* | grep "wp-"
[Fri Feb 18 09:05:49.042574 2022] [php7:error] [pid 3789616] [client 103.109.96.11:60066] script '/var/www/html/eusec20/wp
-login.php' not found or unable to stat
[Thu Feb 17 08:23:31.605082 2022] [php7:error] [pid 3630350] [client 102.165.48.97:40892] script '/var/www/html/wp-login.p
hp' not found or unable to stat
[Thu Feb 17 08:23:31.951171 2022] [php7:error] [pid 3631784] [client 102.165.48.97:40894] script '/var/www/html/eusec20/wp
-login.php' not found or unable to stat
[Thu Feb 17 08:23:31.978838 2022] [php7:error] [pid 3632298] [client 102.165.48.97:40896] script '/var/www/html/eusec/wp-l
ogin.php' not found or unable to stat
[Thu Feb 17 10:03:18.958818 2022] [php7:error] [pid 3641153] [client 47.104.66.61:58626] script '/var/www/html/interestsre
search/wp-login.php' not found or unable to stat, referer: http://interestsresearch.io/wp-login.php
[Thu Feb 17 11:04:27.068009 2022] [php7:error] [pid 3646525] [client 80.251.219.111:60460] script '/var/www/html/computers
ecurityclasscom/wp-login.php' not found or unable to stat, referer: http://computersecurityclass.com/wp-login.php
[Thu Feb 17 11:35:43.470994 2022] [php7:error] [pid 3649892] [client 107.173.165.214:34454] script '/var/www/html/aifairne
sstech/wp-login.php' not found or unable to stat, referer: http://aifairness.tech/wp-login.php
```

Online Tracking

Online Tracking

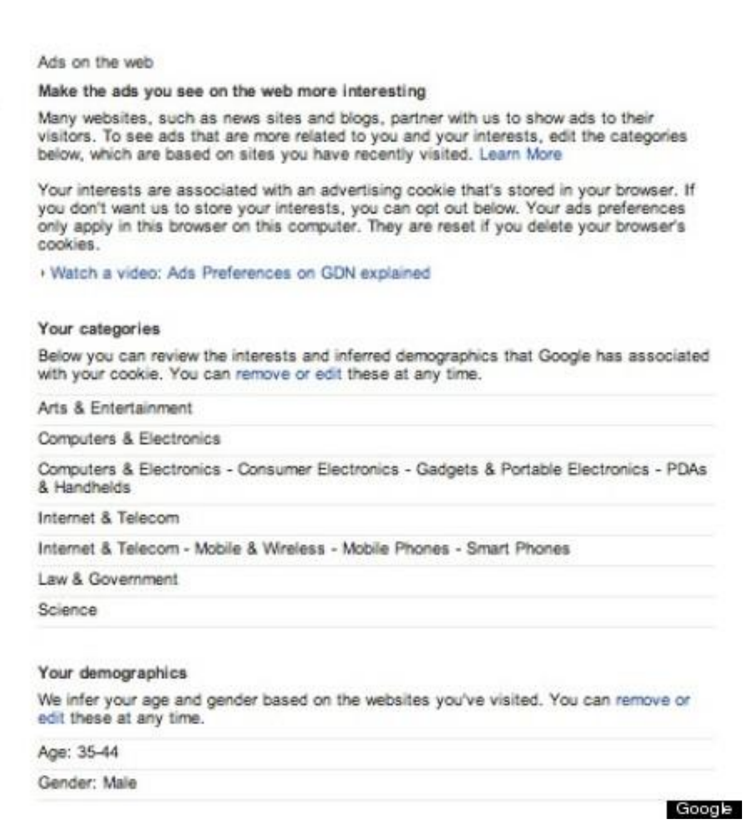
- Advertisers want to show you advertisements targeted to your interests and demographics



How your ads are personalized

Ads are based on personal info you've added to your Google Account, data from advertisers that partner with Google, and Google's estimation of your interests. Choose any factor to learn more or update your preferences. Learn more

Accounting & Finance Jobs	Action & Adventure Films
Action & Platform Games	Adventure Games
Android OS	Autos & Vehicles
Banking	Bars, Clubs & Nightlife
Beaches & Islands	Blues
Bollywood & South Asian Film	Books & Literature
Business & Productivity Software	Business News



Ads Preferences

- Ads on Search and Gmail
- Ads on the web
- Opt out

Ads on the web

Make the ads you see on the web more interesting

Many websites, such as news sites and blogs, partner with us to show ads to their visitors. To see ads that are more related to you and your interests, edit the categories below, which are based on sites you have recently visited. Learn More

Your interests are associated with an advertising cookie that's stored in your browser. If you don't want us to store your interests, you can opt out below. Your ads preferences only apply in this browser on this computer. They are reset if you delete your browser's cookies.

Watch a video: Ads Preferences on GDN explained

Your categories

Below you can review the interests and inferred demographics that Google has associated with your cookie. You can remove or edit these at any time.

Arts & Entertainment

Computers & Electronics

Computers & Electronics - Consumer Electronics - Gadgets & Portable Electronics - PDAs & Handhelds

Internet & Telecom

Internet & Telecom - Mobile & Wireless - Mobile Phones - Smart Phones

Law & Government

Science

Your demographics

We infer your age and gender based on the websites you've visited. You can remove or edit these at any time.

Age: 35-44

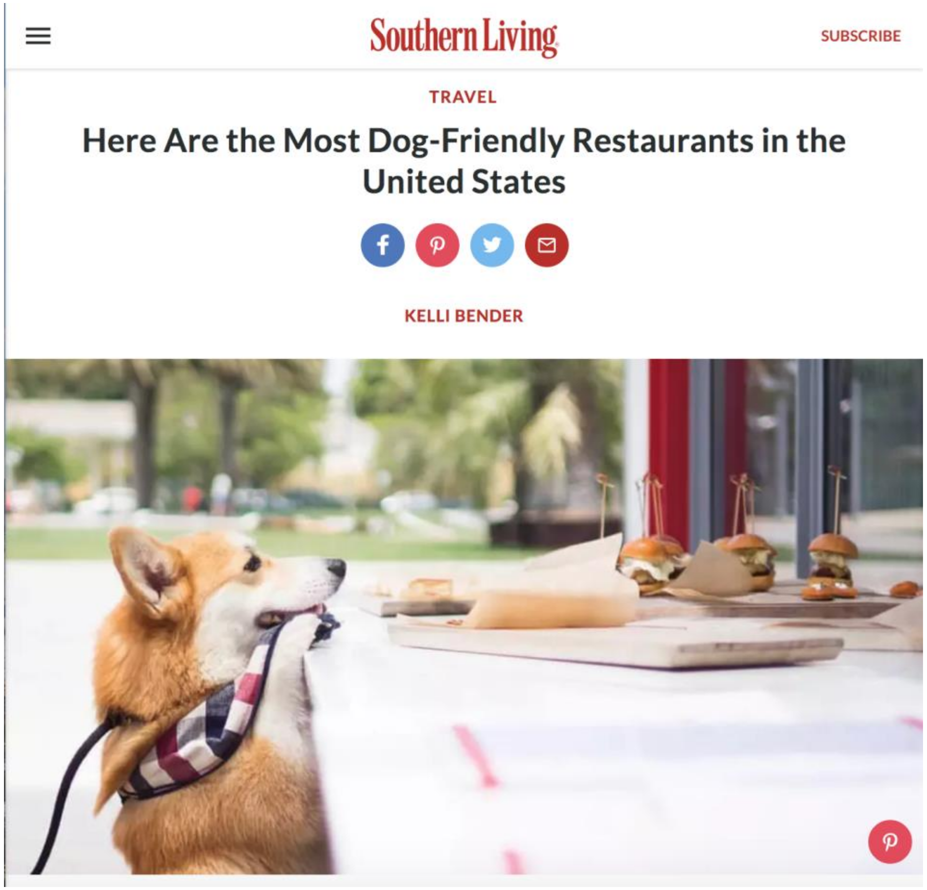
Gender: Male

Google

Online Tracking

- First party = the site you are visiting (whose address is in the URL bar)
- Third party = other sites contacted as a result of your visit to that site
- First-party tracking (e.g., for search)
  - Consider DuckDuckGo and alternatives

Data-Driven Inferences



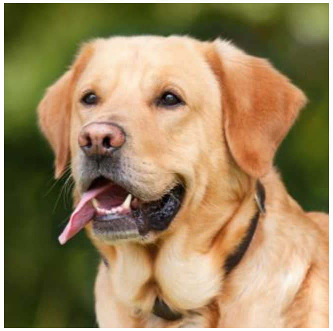
Southern Living

SUBSCRIBE

TRAVEL

Here Are the Most Dog-Friendly Restaurants in the United States

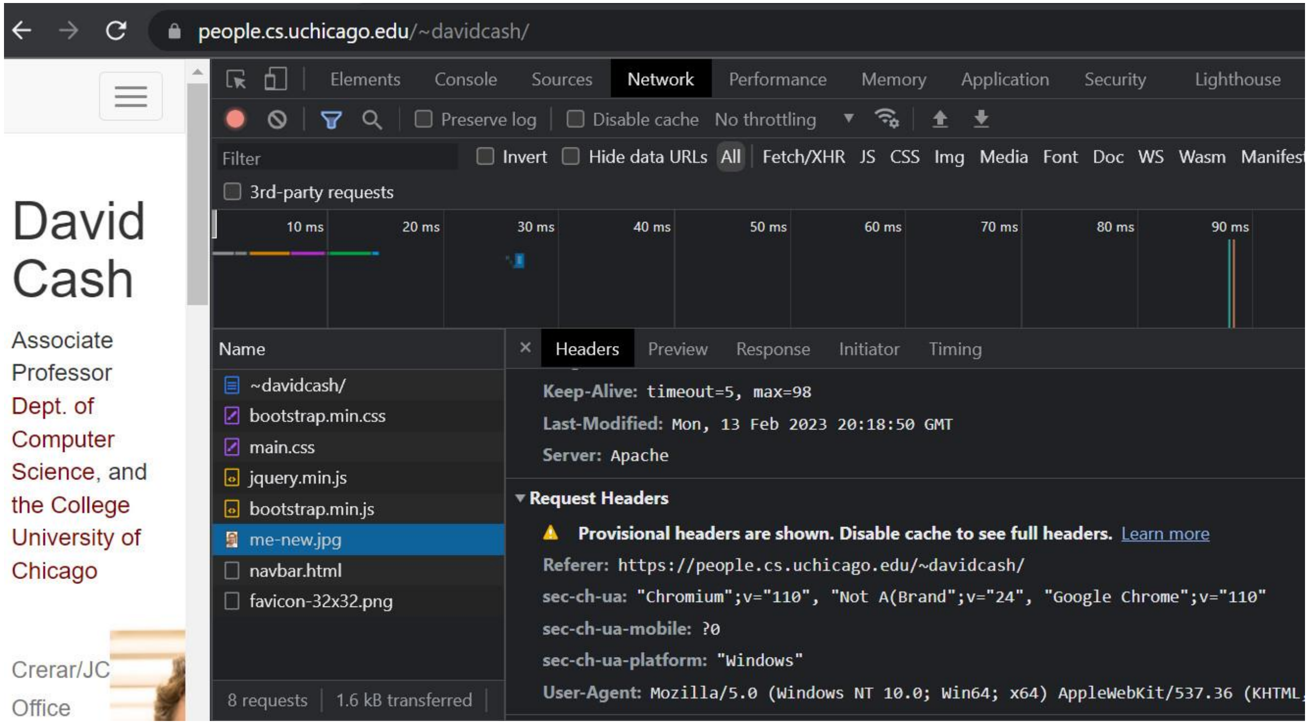
KELLI BENDER



You might like dogs!

Mechanics of Tracking

- Most commonly, tracking is accomplished via HTTP cookies
  - Third-party cookies (+ referrer HTTP header)



```
Keep-Alive: timeout=5, max=98
Last-Modified: Mon, 13 Feb 2023 20:18:50 GMT
Server: Apache
▾ Request Headers
⚠ Provisional headers are shown. Disable cache to see full headers. Learn more
Referer: https://people.cs.uchicago.edu/~davidcash/
sec-ch-ua: "Chromium";v="110", "Not A(Brand";v="24", "Google Chrome";v="110"
sec-ch-ua-mobile: ?0
sec-ch-ua-platform: "Windows"
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML,
```

Mechanics of Online Tracking

- JavaScript / images from advertising networks loaded as part of your page
  - In iframes
  - Or sometimes not
  - Why does this matter?
- Let’s discuss: what can an advertising network learn, and how?

Mechanics of Cookie Syncing

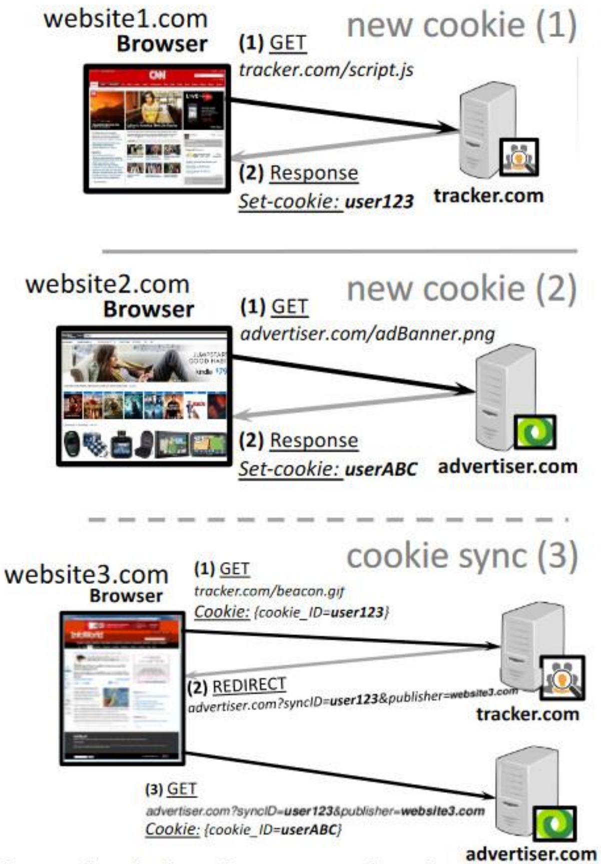


Figure 1: Example of advertiser.com and tracker.com synchronizing their cookieIDs. Interestingly, and without having any code in website3, advertiser.com learns that: (i) cookieIDs userABC==user123 and (ii) userABC has just visited the given website. Finally, both domains can conduct server-to-server user data merges.

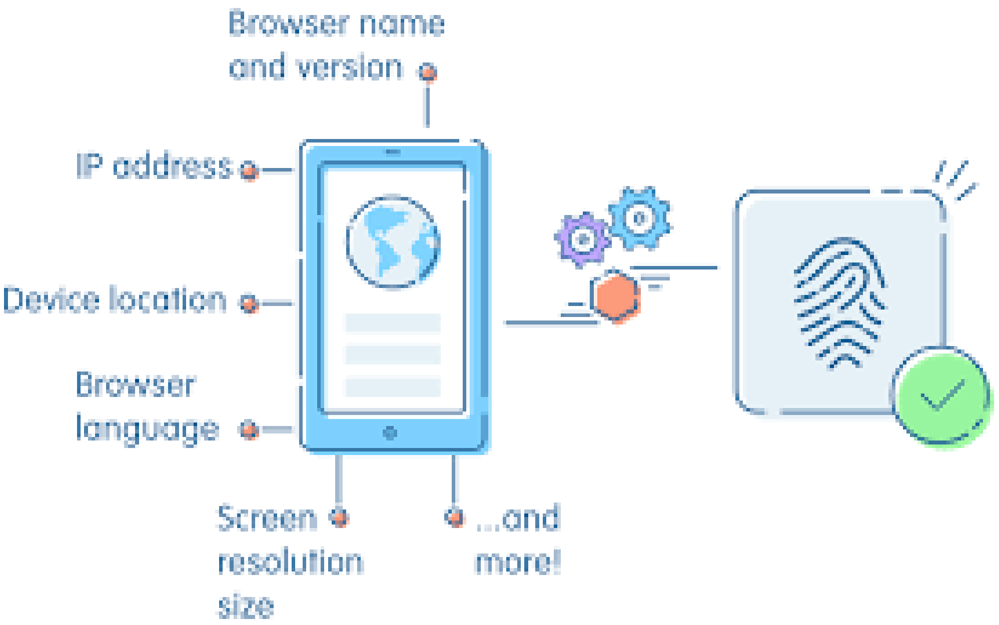
From Papadopoulos et al. “Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask,” in *Proc. WWW*, 2019.

Browser fingerprinting

- Use features of the browser that are relatively unique to your machine
  - Fonts
  - GPU model anti-aliasing (Canvas fingerprinting)
  - User-agent string
  - *(Often not)* IP address *(Why not?)*

Device Fingerprinting

- Use unique(-ish) combination of device features as an identifier
- https://panopticlick.eff.org/

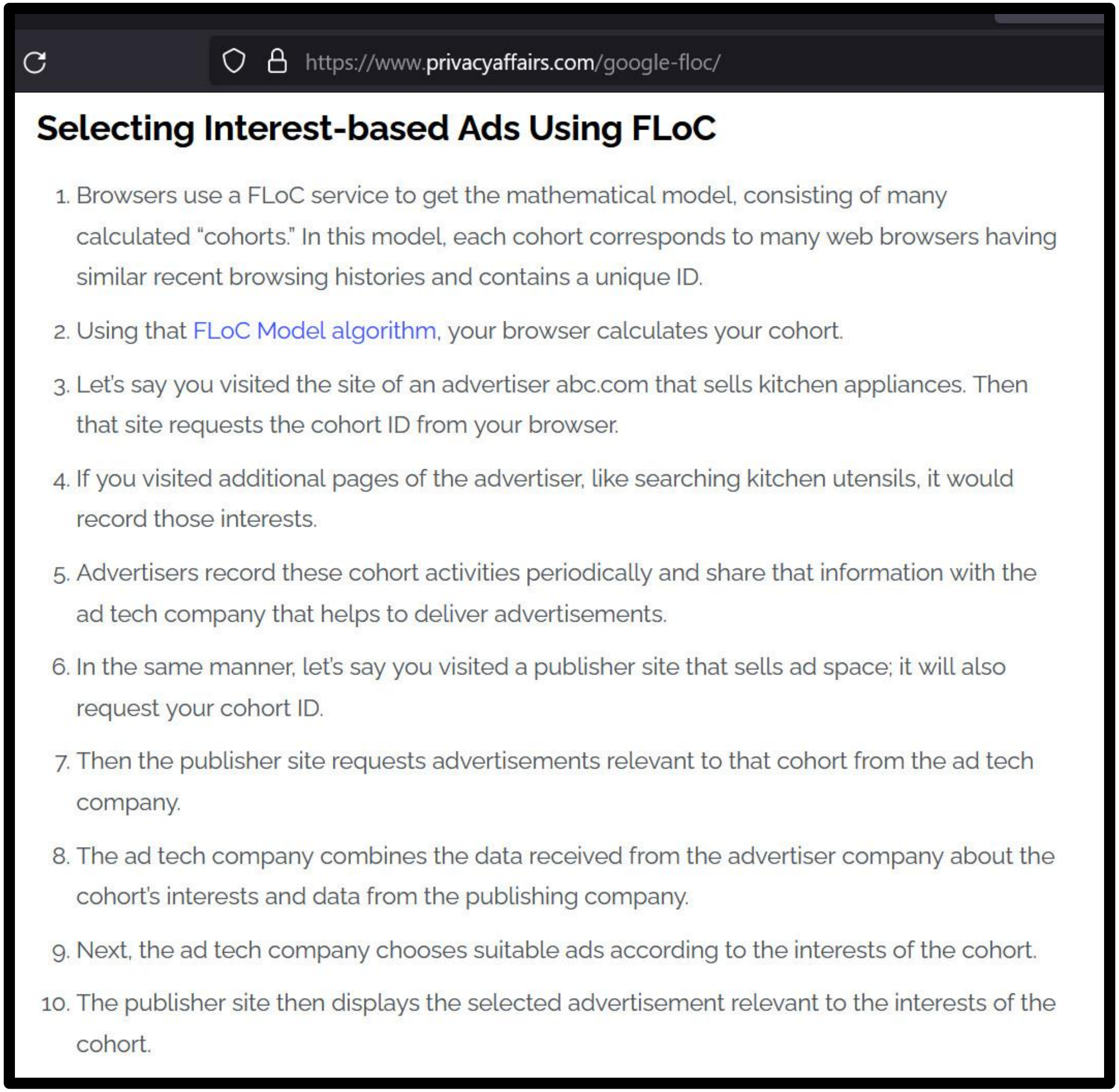


Alternatives to Cookies for Tracking / Profiling

Google's FLoC

- Federated Learning of Cohorts
- Clusters users based on their browsing activity and assigns a cohort ID
  - Uses SimHash for clustering
  - Clusters *intended to* contain 1,000s of users
- Criticisms include fingerprintability, ability to tie cohort to PII, and collapse of different browsing contexts
- (Abandoned in early 2022)

Google’s FLoC

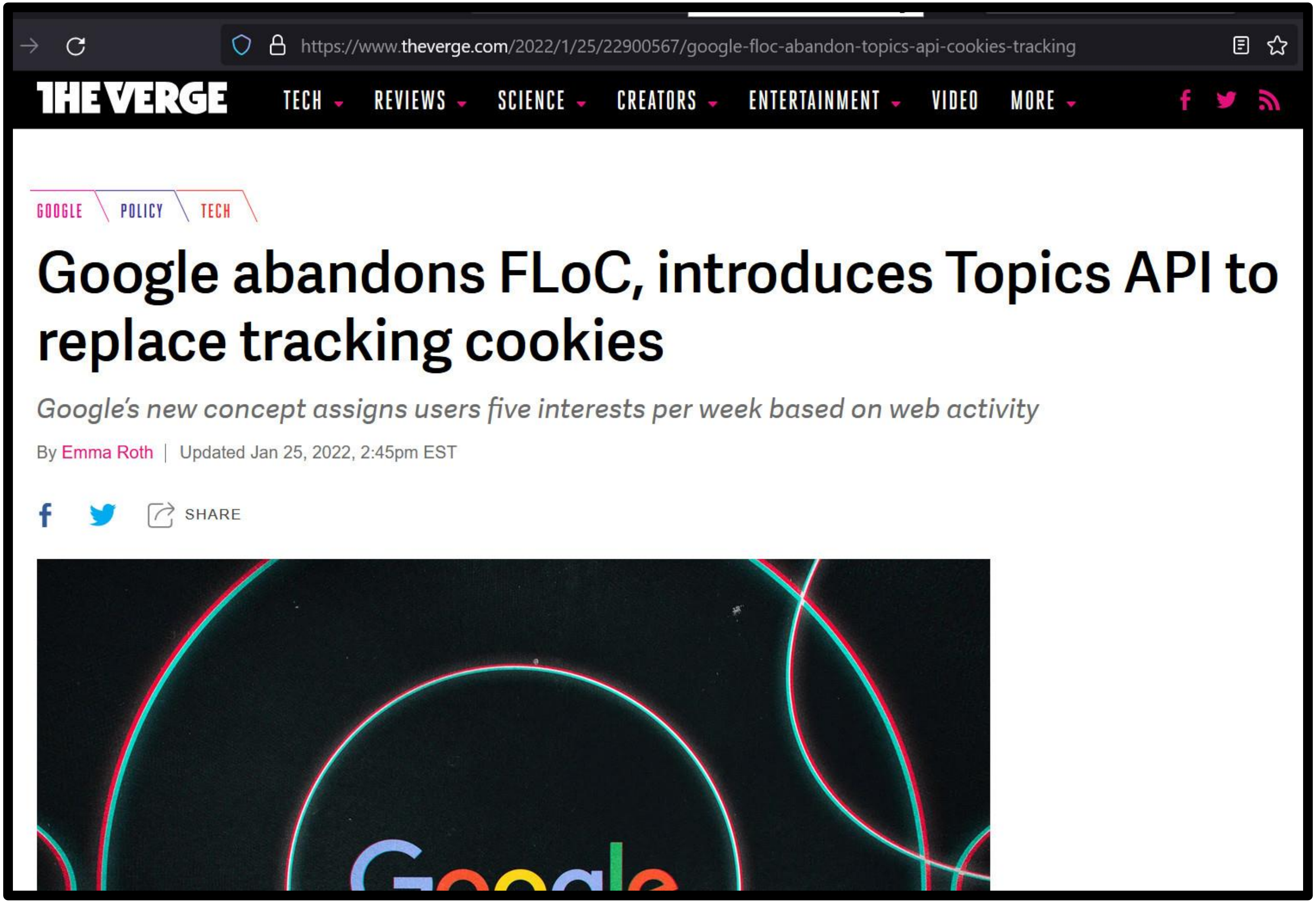


https://www.privacyaffairs.com/google-floc/

Selecting Interest-based Ads Using FLoC

1. Browsers use a FLoC service to get the mathematical model, consisting of many calculated “cohorts.” In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.
2. Using that FLoC Model algorithm, your browser calculates your cohort.
3. Let’s say you visited the site of an advertiser abc.com that sells kitchen appliances. Then that site requests the cohort ID from your browser.
4. If you visited additional pages of the advertiser, like searching kitchen utensils, it would record those interests.
5. Advertisers record these cohort activities periodically and share that information with the ad tech company that helps to deliver advertisements.
6. In the same manner, let’s say you visited a publisher site that sells ad space; it will also request your cohort ID.
7. Then the publisher site requests advertisements relevant to that cohort from the ad tech company.
8. The ad tech company combines the data received from the advertiser company about the cohort’s interests and data from the publishing company.
9. Next, the ad tech company chooses suitable ads according to the interests of the cohort.
10. The publisher site then displays the selected advertisement relevant to the interests of the cohort.

Google’s Topics API

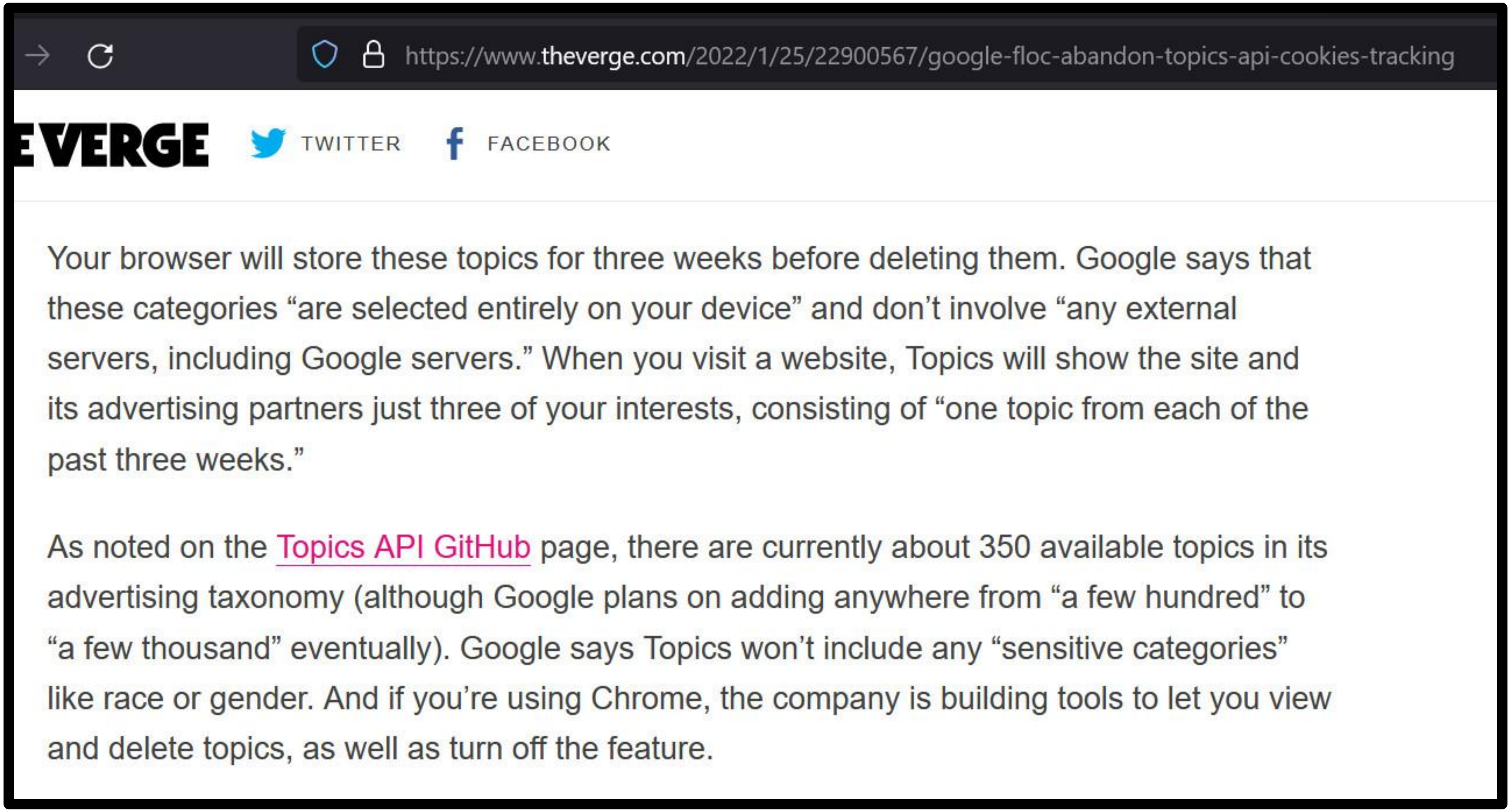

https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking
THE VERGE
TECH
REVIEWS
SCIENCE
CREATORS
ENTERTAINMENT
VIDEO
MORE
GOOGLE
POLICY
TECH
SHARE
Google

Google abandons FLoC, introduces Topics API to replace tracking cookies

Google’s new concept assigns users five interests per week based on web activity

By Emma Roth | Updated Jan 25, 2022, 2:45pm EST

Google’s Topics API



https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking

Your browser will store these topics for three weeks before deleting them. Google says that these categories “are selected entirely on your device” and don’t involve “any external servers, including Google servers.” When you visit a website, Topics will show the site and its advertising partners just three of your interests, consisting of “one topic from each of the past three weeks.”

As noted on the Topics API GitHub page, there are currently about 350 available topics in its advertising taxonomy (although Google plans on adding anywhere from “a few hundred” to “a few thousand” eventually). Google says Topics won’t include any “sensitive categories” like race or gender. And if you’re using Chrome, the company is building tools to let you view and delete topics, as well as turn off the feature.